The complete mtDNA genome of Huaibei Grey donkey: genome characterization and phylogenetic analysis

Jingjing Xia[1], Liang Chang[1], Dashuang Xu[2], Yuqing Jia[1], Yuanfei Ding[1], Chengcheng Cao[1], Zhaoyu Geng[1], and Sihua Jin[1]

[1]Anhui Agricultural University College of Animal Science and Technology
[2]Anhui Livestock and Poultry Genetic Resources Protection Center

August 15, 2022

Abstract

To investigate the conservation and phylogenetic relationship with other breeds of the local Anhui Province Huaibei Grey donkey (HGD), the complete mitochondrial DNA (mtDNA) was sequenced and de novo assembled using deep sequencing data from total genomic DNA. The final size of mtDNA was 16 670 bp (NCBI submission number: MZ911746), including 22 tRNA genes, two rRNA genes, 13 protein-coding genes (PCGs), and one non-coding control region. The PCGs region consisted of 5 559 codons. Most of the PCGs had ATG and TAA as the start and stop codons, respectively. Then we analyzed the proximal part of the D-loop region (418 bp, between 15 419 bp to 15 836 bp) from HGDs, using DNAsp v6 software. We identified 23 polymorphic nucleotide sites and found that the A, C, G, and T bases comprised 30.4%, 34.9%, 13.1%, 21.6%, respectively, of the mtDNA D-loop sequence. The haplotype and nucleotide diversity were 0.87000 and 0.02115, correspondingly. Altogether, the 60 sequences displayed 11 different haplotypes, the most frequent haplotype was H9 (23.33%), followed by H4 (21.67%). MJ network analysis indicated that all haplotypes were clearly divided into Clade I and II lineages, which indicates that HGD may have two maternal lineages. Phylogeographic analysis indicates that the Somali lineage could be the most probable domestication center for HGD. Our study provides an empirical basis for the characterization, conservation, and management of HGD genetic resources.

Running head: Mitochondrial genome and phylogenetic analysis of Huaibei grey donkey

The complete mtDNA genome of Huaibei Grey donkey: genome characterization andphylogenetic analysis

Jingjing Xia[1+] | Liang Chang[1+] | Dashuang Xu[2]| Yuqing Jia[1] | Yunfei Ding[1] | Chengcheng Cao[1]| Zhaoyu Geng[1] | Sihua Jin[1*]

[1] College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China;

[2] Anhui Livestock and Poultry Genetic Resources Protection Center, Hefei 231283, China

[+]These authors contributed equally to this work.

***Correspondence:** Dr. and Associate Prof. Sihua Jin, College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China.

Tel: +86-551-6578 6328

Fax: +86-551-6578 6326

E-mail: jsh3235@126.com

Funding information:

Program for Young Outstanding Scientists of University (gxyqZD2022017)

Anhui Poultry-Industry Technology Research System (AHCYJSTX-06-10)

Municipal Science and Technology Cooperation Project (KJ2021564)

Municipal Science and Technology Cooperation Project (KJ20200323)

The Science and Technology Major Project of Anhui Province (201903a06020018)

Abstract: To investigate the conservation and phylogenetic relationship with other breeds of the local Anhui Province Huaibei Grey donkey (HGD), the complete mitochondrial DNA (mtDNA) was sequenced and de novo assembled using deep sequencing data from total genomic DNA. The final size of mtDNA was 16 670 bp (NCBI submission number: MZ911746), including 22 tRNA genes, two rRNA genes, 13 protein-coding genes (PCGs), and one non-coding control region. The PCGs region consisted of 5 559 codons. Most of the PCGs had ATG and TAA as the start and stop codons, respectively. Then we analyzed the proximal part of the D-loop region (418 bp, between 15 419 bp to 15 836 bp) from HGDs, using DNAsp v6 software. We identified 23 polymorphic nucleotide sites and found that the A, C, G, and T bases comprised 30.4%, 34.9%, 13.1%, 21.6%, respectively, of the mtDNA D-loop sequence. The haplotype and nucleotide diversity were 0.87000 and 0.02115, correspondingly. Altogether, the 60 sequences displayed 11 different haplotypes, the most frequent haplotype was H9 (23.33%), followed by H4 (21.67%). MJ network analysis indicated that all haplotypes were clearly divided into Clade I and II lineages, which indicates that HGD may have two maternal lineages. Phylogeographic analysis indicates that the Somali lineage could be the most probable domestication center for HGD. Our study provides an empirical basis for the characterization, conservation, and management of HGD genetic resources.

KEYWORDS: Huaibei grey donkey, mtDNA, D-loop, haplotype, genetic diversity, maternal origin

1 | INTRODUCTION

Mitochondrial DNA (mtDNA) is an extranuclear genetic material with a specific structure that does not undergo recombination during generational transmission and is maternally derived (Xia et al., 2019). As a genetic marker, mtDNA is of great significance in genetic and evolutionary studies of livestock (Bruford et al., 2003; Saccone et al., 2000). mtDNA contains exons and noncoding regions - the control region (CR) - also known as the displacement-loop region (D-loop) (Wolf et al., 1999). In the mammalian mtDNA D-loop, the most variable region exists between $tRNA^{Pro}$ and the large conserved sequence block (Nobushige et al., 1995). Additionally, mtDNA has been used to infer the wild ancestors and determine domestication centers of modern domestic animals, and to study the origins, evolutionary relationships and genetic diversity of several domestic animals (Bruford et al., 2003; Gissi et al., 2000).

The domestication of donkey is generally considered to have taken place in the tropics or subtropics of Africa (Beja-Pereira et al., 2004), though it remains unclear. Studies of donkey mitochondrial sequences have suggested that there were two highly differentiated maternal lineages (Clade I and Clade II) during donkey domestication (*Equus asinus*), Clade I clustered clearly with the Nubian wild ass (*E. africanus africanus*); Clade II originated from the Somali wild ass (*E. africanus somaliensis*) which was probably on the verge of extinction (Beja-Pereira et al., 2004; Chen et al., 2006; Kimura et al., 2011). In recent years, an increasing number of studies on donkeys' mtDNA have assessed the evolutionary relationship, variation and genetic diversity among Asian, European, and African breeds (Gan et al., 2011; Han et al., 2014; Lei et al., 2007; Ma et al., 2020; Ozkan Unal et al., 2020; Perez-Pardal et al., 2014).

The three groups of Chinese donkey are distinguished based on body size: large, medium, and small donkeys. Huaibei grey donkey (HGD) is a small donkeys with grey fur, and its main area of production is in Huaibei City, Anhui Province. Few reports have analyzed mtDNA D-loop polymorphisms in HGD. Therefore, in the present study we used PCR sequencing to detect partial sequence polymorphisms of the D-loop and analyzed the mtDNA diversity and evolutionary relationships of HGD to inform its management.

2 | MATERIALS AND METHODS

2.1 | Ethical statement

Samples were collected from HGDs at Anhui Domestic Donkey Conservation and Breeding, Anhui, China (Figure 1). All procedures were approved by the Animal Care and Use Committee of Anhui Agricultural University (SYXK 2016-007).

2.2 | Sample collection and DNA isolation

Blood samples were obtained and genomic DNA was isolated using a blood DNA extraction kit (TIANGEN, Beijing, China) and then stored at -20 . DNA integrity was evaluated by 1.5% agarose gel electrophoresis. The most variable region in the mtDNA D-loop was assessed using data for each of the samples. One female HGD was randomly selected to determine the complete mtDNA sequence.

2.3 | Amplification part of mtDNA control regions (D-Loop) and sequencing

The proximal part of the D-loop region (418 bp, between 15 419 bp to 15 836 bp, GenBank accession number X97337(Xu et al., 1996)) between the*trnP* gene and the central conserved sequence block (Ishida et al., 1994) , was amplified by PCR using appropriate primers (F: 5'-ACCACTCGCAAGCACCA-3'; R: 5'-CACAGCATCCCCAAATA-3') as previously described (Ivankovic et al., 2002). The primers were designed by Primer3 (Untergasser et al., 2012) and synthesized and purified by TSINGKE Biological Technology (Nanjing, China). PCR amplification was performed in a 50 μL system containing 300-500 ng of template DNA, 25 μL 2 $\times$ Taq Master Mix, 2 μL of each primer, and 19 μL of ddH_2O. Amplification conditions were as follows: 94 °C for 5 min, 35 cycles at 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension step at 72 °C for 10 min. The complete mtDNA of one HGD was obtained by whole-genome shotgun sequencing, using an Illumina NovaSeq 6000 platform, with paired-end read lengths of 150 bp.

2.4 | Sequence assembly and sequence analysis

After quality control, clean data from complete mtDNA were aligned against a reference genome (GenBank, NC_001788.1) using Bowtie2. Reads on the calibration were reserved and SPAdes v3.13.0 (Parameters: K 127) was used for genome assembly. The splicing result was compared with the closed reference genome using BLASTN, and the assembly result was determined accordingly. tRNAs were recognized and their secondary structures were predicted by tRNAscan-SE v2.0 (Chan & Lowe, 2019). The most variable region was identified by sequence alignment against the reference genome. We derived a circular map of the mitogenome using ORDRAW (Greiner et al., 2019), and base composition skew was calculated using AT-skew = (A - T)/(A + T) and GC-skew = (G - C)/(G + C) (Perna & Kocher, 1995).

The 418 bp section was manually edited and aligned against the reference sequence using CLUSTALX v2.0 (Larkin et al., 2007). DNAsp v6 (Rozas et al., 2017) was used to calculate the nucleotide diversity, polymorphic sites, and haplotype diversity of the most variable region in the HGD mtDNA D-loop region. Genetic relationships among populations were determined using median-joining networks (MJN) using Network v.10.1.0.0 software (Bandelt et al., 1999), the MJN was constructed from 11 haplotypes found in this study and 60 reference sequences found previously (Ozkan Unal et al., 2020; Xia et al., 2019), the 15 sequences including Hap4, Hap6, Hap7, and Hap10, Hap12, Hap16, Hap18, Hap20, Hap22, Hap23, Hap24, Hap25, Hap26, Hap27, Hap29 belonged to Clade I. Reference mtDNA D-loop region sequences of wild and domestic donkey were obtained from GenBank and used as comparators of HGDs to build phylogenetic trees and speculate their origin (Han et al., 2014; Ivankovic et al., 2002; Kimura et al., 2011; Oakenfull et al., 2000; Ozkan Unal et al., 2020). The phylogenetic tree was drawn using the Neighbor-joining (NJ) method. Bootstrap values were estimated using 1,000 repetitions (Felsenstein, 1985), based on the Kimura-2-parameter genetic distances (Saitou & Nei, 1987), and reconstructed using MEGA v7.0 (Kumar et al., 2016).

3 | RESULTS

3.1 | Structure and organization of complete mtDNA

We deposited the complete mtDNA - a circular 16 670 bp molecule - in GenBank under accession no. MZ911746. The gene map of Huaibei grey donkey mtDNA is presented in Figure 2. There were 13 protein coding genes (PCGs) including 7 NADH dehydrogenase complex subunits (*ND1-6* and *ND4L*), three cytochrome oxidase subunits (*COX1-3*), cytochrome b (*CYTB*) and two ATPase subunits (*ATP6* and *ATP8*). The mitogenome also contains two rRNA genes (*rRNAL* , *rRNAS*), 22tRNA genes and a control region (D-loop). With respect to their location, 14 tRNA genes, 12 PCGs and two rRNA genes were located on the positive strand and the remaining eight tRNA genes and one PCG on the negative strand (Table 1). On the whole, a bias in nucleotide composition was observed toward A (32.3%) and T (25.6%), with respect to C (28.9%) and G (13.2%). We observed a negative GC-skew (–0.3729) and a positive AT-skew (+0.1157), indicating that A and C were marginally more numerous than T and G (Table 2).

3.2 | Condon usage and protein coding genes

The total length of the 13 protein coding genes (PCGs) was 9 911 bp, and they contained 57.3% AT nucleotides, with a positive AT-skew (+0.0750) and a negative GC-skew (–0.4239). Relative synonymous codon usage values for the Huaibei grey donkey were shown in Table 3. Codons encoding Trp were infrequent, while those encoding Leu and Ser occur most frequently (Figure 3). The PCGs region consisted of a total of 5 559 codons. The initiation and termination signals, as well as gene lengths, are listed in Table 2. Eleven PCGs had the ATG start codon, with the exception for *ND2*and *ND3* , which had ATA as the start codon. Eight PCGs (*ND1* , *COX1* , *COX2* , *ATPase8* , *ATPase6* ,*ND4L* , *ND5* , and *ND6*) had TAA as the stop codon;*ND2* , *COX3* , and *ND3* had TAG as the stop codon; and*ND4* and *CYTB* had AGA as the stop codon. The most frequently used amino acids by the PCGs of the mitogenomes of Huaibei grey donkey were leucine and serine (12.4% and 12.4%); among the 64 available codons, the three most frequently used codons are CUA (2.12%) for Leu2, UCA (3.18%) for Ser1, and GGA (2.41%) for Gly.

3.3 | Transfer and ribosomal RNA genes

The total length of the 22 tRNAs in Huaibei grey donkey mtDNA was 1 516 bp with an A + T content of 61.7%, positive AT-skew (+0.1183), and negative GC-skew (–0.1854). Most tRNA genes had the common cloverleaf secondary structure, excluding tRNA-Ser (GCT), which lacked the dihydrouridine arm (Figure 4). The total size of the two rRNAs was 2 555 bp, the A + T content was 60.1%, the AT-skew was positive (+0.2164), and the GC-skew was negative (–0.1508). The *rrnL* was located between *trnV* and *trnL1* , and the *rrnS* was located between *trnV* and *trnF* . The length of the *rrnL* and*rrnS* was 1 580 bp and 975 bp, respectively.

3.4 | Genetic variation and genetic diversity of complete D-loop region

The mtDNA D-loop sequences had a total of 23 variable sites and 11 haplotypes (named sequentially from H1 to H11, GenBank accession no. OP095367˜OP095377.) (Table 4). The A + T content of D-loop region was 52%, with a positive AT-skew (+0.1692) and negative GC-skew (–0.4542), which was found between*trnP* and *trnF* . The haplotype and nucleotide diversity values were 0.87000 +- 0.00046 and 0.02115 +- 0.01180, respectively. There were two single polymorphic sites and 21 parsimony-informative polymorphic sites. All polymorphisms were A/G and T/C transitions. We observed several variable sites in Clade II (66 A/G, 72 C/T, 85 C/T, 151 A/G, 162 A/G, 180 T/C, 226 A/G, 234 T/C, 224 A/G, 280 T/C, 352 T/C, 383 T/C, 388 T/C, 402 T/C, and 404 A/G) and only two polymorphic sites in Clade I(181 A/G and 403 A/G). The calculated number of polymorphic sites and haplotypic diversity of Huaibei grey donkey are shown in Table 4.

3.5 | Maternal origin of Huaibei grey donkey

The molecular studies of the donkey mitochondrial sequence have apparently defined two distinct matriarchal (Clade I and II lineages) related to domestication. Comparison of the 60 reference mtDNA D-loop sequences (Table 5) and the HGD populations sequences of 418 bp were made to determine the relationships among the haplotypes and population structure of Huaibei grey donkey; the median joining network was constructed for the identified haplotypes. There are two distinct lineages (Clade I and Clade II) as shown in the median

joining network, and most of the HGDs were classified into Clade I (31 individuals [51.67%] and 4 haplotypes), whereas Clade II included 29 individuals (48.33%) comprising 7 haplotypes (Figure 5).

To further clarify the origin of HGDs, we used mtDNA sequence comparisons with the Nubian wild ass (*E. africanus africanus*), the Somali wild ass (*E. africanus somaliensis*), Asian wild ass (*E. hemionus*) , European and Chinese domestic donkeys. HGDs was clearly clustered within the Somali wild ass (*E. africanus somaliensis*) sequences. They were clustered apart from the European and Chinese domestic donkey clade, as indicated in the phylogenetic tree, therefore these results indicate that Africa was the most probable location for donkey domestication (Figure 6).

4 | DISCUSSION

The mtDNA genome of donkeys is 16 670 bp, of which the D-loop region is 1 207 bp (Xu et al., 1996). As reported in previous genetic studies on donkeys (Lopez et al., 2005; Ma et al., 2020; Ozkan Unal et al., 2020), the D-loop region of ancestral mtDNA supplies adequate information to assess genetic variation, evolutionary relationships, and matrilineal genetic origins.

We detected a total of 23 nucleotide polymorphic sites in the D-loop region sequences of HGDs. For the D-loop 418 bp sequence, nucleotide polymorphic sites account for 9.21%, among which 94.44% are transformed or transposed, with two insertion sites accounting for 5.56% (Ivankovic et al., 2002). Therefore, the frequency of transposition was much higher than transversion, which is consistent with the results of our study.

At present, there is a traditional view on the origin and evolution of the Chinese donkey, which is that the Chinese donkey originated from the African wild ass after domestication. Lei et al. (Lei et al., 2007) found two mitochondrial origins of African wild ass, lineage Somali (Clade II) and lineage Nubian (Clade I) in Chinese domestic donkeys and that Clade II was prevalent in Chinese domestic donkey breeds. These findings are consistent with those of our study.

5 | CONCLUSION

The present study demonstrated the abundant mtDNA diversity existing in HGDs, as indicated by the presence of 11 haplotypes. In summary, our results showed that the mtDNA D-loop region of HGDs had a high genetic diversity, we confirmed that Chinese donkey originated from the African wild ass after domestication, as reported in previous research, and we could clearly exclude the Asian wild ass as an ancestor of the HGD. The study has provided an empirical basis for the elucidation of the genetic structure of the maternal ancestor of Huaibei grey donkeys.

ACKNOWLEDGEMENTS

We thanked Anhui Domestic Donkey Conservation and Breeding for providing samples. The authors thank the anonymous referees for their valuable comments and suggestions on the manuscript. This study was supported by the projects Program for Young Outstanding Scientists of University (gxyqZD2022017); Anhui Poultry-Industry Technology Research System (AHCYJSTX-06-10); Municipal Science and Technology Cooperation Project (KJ2021564); Municipal Science and Technology Cooperation Project (KJ20200323); The Science and Technology Major Project of Anhui Province(201903a06020018).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data presented in this study are openly available in NCBI GenBank with accession number MZ911746 for Huaibei grey donkey's complete mitochondrial DNA sequence, and the accession numbers OP095367˜OP095377 for Huaibei grey donkey's 11 haplotype sequences.

AUTHOR CONTRIBUTION

Jingjing Xia and Liang Chang designed the work; Jingjing Xia, Yuqing Jia, Dashuang Xu, Zhaoyu Geng and Sihua Jin collected samples and contributed genetic data; Jingjing Xia, Liang Chang, Yuanfei Ding and Chengcheng Cao contributed data analysis; Jingjing Xia and Liang Chang drafted the manuscript. All authors critically revised and approved the final manuscript.

ORCID

Jingjing Xia https://orcid.org/0000-0002-9760-2782

Liang Chang https://orcid.org/0000-0002-7917-9984

Dashuang Xu https://orcid.org/0000-0002-8567-7584

Yuqing Jia https://orcid.org/0000-0001-7013-992X

Yunfei Ding https://orcid.org/0000-0003-3305-1560

Chengcheng Cao https://orcid.org/0000-0002-3475-2716

Zhaoyu Geng https://orcid.org/0000-0003-1898-7459

Sihua Jin https://orcid.org/0000-0002-0212-2701

REFERENCES

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* , 1(16), 37-48.

Beja-Pereira, A., England, P. R., Ferrand, N., Jordan, S., Bakhiet, A. O., Abdalla, M. A., Mashkour, M., Jordana, J., Taberlet, P., & Luikart, G. (2004). African origins of the domestic donkey. *Science* , 5678(304), 1781.

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* , 11(4), 900-910.

Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* , 1962), 1-14.

Chen, S. Y., Zhou, F., Xiao, H., Sha, T., Wu, S. F., & Zhang, Y. P. (2006). Mitochondrial DNA diversity and population structure of four Chinese donkey breeds. *Animal Genetics* , 4(37), 427-429.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* , 4(39), 783-791.

Gan, Q. F., Lin, S. X., Liang, X. W., Meng, H. L., Qiao, H., Du, Y. F., Ke, F. R., & Zheng, S. K. (2011). African maternal origin and genetic diversity of five chinese domestic donkeys breeds. *Journal of Animal and Veterinary Advances* , 23(10), 3090-3094.

Gissi, C., Reyes, A., Pesole, G., & Saccone, C. (2000). Lineage-specific evolutionary rate in mammalian mtDNA. *Molecular Biology and Evolution* , 7(17), 1022-1031.

Greiner, S., Lehwark, P., & Bock, R. (2019). Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* , W1(47), W59-W64.

Han, L., Zhu, S., Ning, C., Cai, D., Wang, K., Chen, Q., Hu, S., Yang, J., Shao, J., Zhu, H., & Zhou, H. (2014). Ancient DNA provides new insight into the maternal lineages and domestication of Chinese donkeys.*BMC Evolutionary Biology* , 14), 246.

Ishida, N., Hasegawa, T., Takeda, K., Sakagami, M., Onishi, A., Inumaru, S., Komatsu, M., & Mukoyama, H. (1994). Polymorphic sequence in the D-loop region of equine mitochondrial DNA. *Animal Genetics* , 4(25), 215-221.

Ivankovic, A., Kavar, T., Caput, P., Mioc, B., Pavic, V., & Dovc, P. (2002). Genetic diversity of three donkey populations in the Croatian coastal region. *Animal Genetics* , 3(33), 169-177.

Kimura, B., Marshall, F. B., Chen, S., Rosenbom, S., Moehlman, P. D., Tuross, N., Sabin, R. C., Peters, J., Barich, B., Yohannes, H., Kebede, F., Teclai, R., Beja-Pereira, A., & Mulligan, C. J. (2011). Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B-Biological Sciences* , 1702(278), 50-57.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.*Molecular Biology and Evolution* , 7(33), 1870-1874.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* , 21(23), 2947-2948.

Lei, C. Z., Ge, Q. L., Zhang, H. C., Liu, R. Y., Zhang, W., JIang, Y. Q., Dang, R. H., Zheng, H. L., Hou, W. T., & Chen, H. (2007). African maternal origin and genetic diversity of chinese domestic donkeys.*Asian-Australasian Journal of Animal Sciences* , 5(20), 645-652.

Lopez, C. L., Alonso, R., & Aluja, A. S. D. (2005). Study of the genetic origin of the Mexican creole donkey (*Equus asinus*) by means of the analysis of the D-loop region of mitochondrial DNA. *Tropical Animal Health and Production* , 37), 173-188.

Ma, X. Y., Ning, T., Adeola, A. C., Li, J., Esmailizadeh, A., Lichoti, J. K., Agwanda, B. R., Isakova, J., Aldashev, A. A., Wu, S. F., Liu, H. Q., Abdulloevich, N. T., Afanasevna, M. E., Ibrohimovich, K. B., Adedokun, R. A. M., Olaogun, S. C., Sanke, O. J., Mangbon, G. F., Chen, X., Yang, W. K., Wang, Z., Peng, M. S., Ommeh, S. C., Li, Y., & Zhang, Y. P. (2020). Potential dual expansion of domesticated donkeys revealed by worldwide analysis on mitochondrial sequences. *Zoological Research* , 1(41), 51-60.

Nobushige, I., Tsendsuren, O., Suguru, M., Harutaka, M., & Naruya, S. (1995). Mitochondrial DNA sequences of various species of the genus*Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. *Journal of Molecular Evolution* , 2(41), 180-188.

Oakenfull, E. A., Lim, H. N., & Ryder, O. A. (2000). A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus* . *Conservation Genetics* , 4(1), 341-355.

Ozkan Unal, E., Ozdil, F., Kaplan, S., Gurcan, E. K., Genc, S., Arat, S., & Soysal, M. I. (2020). Phylogenetic relationships of Turkish indigenous donkey populations determined by mitochondrial DNA D-loop region. *Animals (Basel)* , 11(10), 1970-1991.

Perez-Pardal, L., Grizelj, J., Traore, A., Cubric-Curik, V., Arsenos, G., Dovenski, T., Markovic, B., Fernandez, I., Cuervo, M., Alvarez, I., Beja-Pereira, A., Curik, I., & Goyache, F. (2014). Lack of mitochondrial DNA structure in Balkan donkey is consistent with a quick spread of the species after domestication. *Animal Genetics* , 1(45), 144-147.

Perna, N. T., & Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* , 3(41), 353-358.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* , 12(34), 3299-3302.

Saccone, C., Gissi, C., Lanave, C., Larizza, A., Pesole, G., & Reyes, A. (2000). Evolution of the mitochondrial genetic system: an overview.*Gene* , 1(261), 153-159.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* , 4(4), 406-425.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Research* , 15(40), e115.

Wolf, C., Rentsch, J., & Hubner, P. (1999). PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification.*Journal of Agricultural and Food Chemistry* , 4(47), 1350-1355.

Xia, X., Yu, J., Zhao, X., Yao, Y., Zeng, L., Ahmed, Z., Shen, S., Dang, R., & Lei, C. (2019). Genetic diversity and maternal origin of Northeast African and South American donkey populations. *Animal Genetics* , 3(50), 266-270.

Xu, X., Gullberg, A., & Arnason, U. (1996). The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution* , 5(43), 438-446.

Lists of Tables and Figures:

TABLE 1 Summary of the Huaibei Grey donkey mitogenome

Gene name	Begin and end	Length/bp	Strand	Anticodon	Start /Stop codon
tRNA-Phe	1˜71	71	+	GAA	-
s-rRNA	72˜1046	975	+	-	-
tRNA-Val	1046˜1112	67	+	UAC	-
1-rRNA	1113˜2692	1580	+	-	-
tRNA-Leu	2693˜2767	75	+	UAA	-
ND1	2770˜3726	957	+	-	ATG/TAA
tRNA-Ile	3726˜3794	69	+	GAU	-
tRNA-Gln	3792˜3864	73	-	UUG	-
tRNA-Met	3867˜3935	69	+	CAU	-
ND2	3936˜4976	1041	+	-	ATA/TAG
tRNA-Trp	4975˜5043	69	+	UCA	-
tRNA-Ala	5049˜5117	69	-	UGC	-
tRNA-Asn	5119˜5191	73	-	GUU	-

Gene name	Begin and end	Length/bp	Strand	Anticodon	Start /Stop codon
tRNA-Cys	5224˜5289	66	-	GCA	-
tRNA-Tyr	5290˜5356	67	-	GUA	-
COX1	5358˜6902	1545	+	-	ATG/TAA
tRNA-Ser	6900˜6968	69	-	UGA	-
tRNA-Asp	6977˜7043	67	+	GUC	-
COX2	7045˜7728	684	+	-	ATG/TAA
tRNA-Lys	7732˜7800	69	+	UUU	-
ATP8	7802˜8005	204	+	-	ATG/TAA
ATP6	7963˜8643	681	+	-	ATG/TAA
COX3	8643˜9427	785	+	-	ATG/TAG
tRNA-Gly	9427˜9496	70	+	UCC	-
ND3	9497˜9843	347	+	-	ATA/TAG
tRNA-Arg	9844˜9912	69	+	UCG	-
ND4L	9914˜10210	297	+	-	ATG.TAA
ND4	10204˜11581	1378	+	-	ATG/AGA
tRNA-His	11582˜11650	69	+	GUG	-
tRNA-Ser	11651˜11710	60	+	UGA	-
tRNA-Leu	11712˜11781	70	+	UAG	-
ND5	11773˜13602	1830	+	-	ATA/TAA
ND6	13586˜14110	525	-	-	ATG/TAA
tRNA-Glu	14114˜14182	69	-	UUC	-
Cytb	14187˜15326	1140	+	-	ATG/AGA
tRNA-Thr	15327˜15398	72	+	UGU	-
tRNA-Pro	15400˜15465	66	-	UGG	-
D-loop	15466˜16681	1216	+	-	-

TABLE 2 Composition and skew values for the Huaibei Grey donkey mitogenome

Type	Size(bp)	A%	T%	G%	C%	AT%	GC%	AT-skew	GC-skew
Complete mitogenome	16680	32.3	25.6	13.2	28.9	57.9	42.1	0.1157	-0.3729
PCGs	9911	30.8	26.5	12.3	30.4	57.3	42.7	0.0750	-0.4239
tRNAs	1516	34.5	27.2	15.6	22.7	61.7	38.3	0.1183	-0.1854
rRNAs	2555	36.5	23.6	16.9	22.9	60.1	39.8	0.2146	-0.1508
Control region	1215	30.4	21.6	13.1	34.9	52	48	0.1692	-0.4542

TABLE 3 Codon number and relative synonymous codon usage (RSCU) of Huaibei Grey donkey mitochondrial protein-coding genes (PCGs)

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	47	0.65	UCU(S)	65	1.01	UAU(Y)	73	0.94	UGU(C)	9	0.82
UUC(F)	98	1.35	UCC(S)	108	1.67	UAC(Y)	82	1.06	UGC(C)	13	1.18
UUA(L)	73	1.13	UCA(S)	126	1.95	UAA(*)	63	1.06	UGA(*)	64	1.07
UUG(L)	17	0.26	UCG(S)	34	0.53	UAG(*)	52	0.87	UGG(W)	9	1
CUU(L)	53	0.82	CCU(P)	75	1.08	CAU(H)	72	0.98	CGU(R)	9	0.71
CUC(L)	75	1.16	CCC(P)	84	1.21	CAC(H)	75	1.02	CGC(R)	14	1.11
CUA(L)	137	2.12	CCA(P)	110	1.58	CAA(Q)	76	1.27	CGA(R)	21	1.66
CUG(L)	32	0.5	CCG(P)	9	0.13	CAG(Q)	44	0.73	CGG(R)	10	0.79

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
AUU(I)	62	0.65	ACU(T)	77	1.04	AAU(N)	62	0.89	AGU(S)	14	0.22
AUC(I)	125	1.3	ACC(T)	80	1.08	AAC(N)	77	1.11	AGC(S)	40	0.62
AUA(I)	101	1.05	ACA(T)	110	1.49	AAA(K)	90	1.75	AGA(R)	10	0.79
AUG(M)	36	1	ACG(T)	29	0.39	AAG(K)	13	0.25	AGG(R)	12	0.95
GUU(V)	22	0.86	GCU(A)	46	1.07	GAU(D)	31	0.93	GGU(G)	14	0.5
GUC(V)	28	1.1	GCC(A)	65	1.51	GAC(D)	36	1.07	GGC(G)	36	1.27
GUA(V)	38	1.49	GCA(A)	53	1.23	GAA(E)	64	1.56	GGA(G)	52	1.84
GUG(V)	14	0.55	GCG(A)	8	0.19	GAG(E)	18	0.44	GGG(G)	11	0.39

TABLE 4 Haplotypes and polymorphic sites of the Huaibei Grey donkey mitogenome

Haplotypes (individual number)	Haplogroup	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites	Localization of polymorphic sites
		66	72	85	151	162	172	174	180	181	202	203	226	227	234	244	280
X97337		G	C	T	A	A	A	A	C	A	T	A	G	A	C	A	C
H1	Clade II	A	T	C	G	G	*	*	T	G	C	G	A	*	T	G	T
H2	Clade I	*	*	*	*	*	*	*	*	G	*	*	*	*	*	*	*
H3	Clade II	A	T	C	G	*	*	*	T	G	*	*	A	G	T	G	T
H4	Clade II	A	T	C	G	G	*	*	T	G	*	*	A	*	T	G	T
H5	Clade I	*	*	*	*	*	*	*	*	G	*	*	*	*	*	*	*
H6	Clade II	A	T	C	G	G	*	*	T	G	*	*	A	G	T	G	T
H7	Clade II	A	T	C	G	*	*	*	T	G	*	*	A	*	T	G	T
H8	Clade II	A	T	C	G	G	*	G	T	G	*	*	A	G	T	G	T
H9	Clade I	*	*	*	*	*	*	*	*	G	*	*	*	*	*	*	*
H10	Clade II	A	T	C	G	G	G	*	T	G	*	*	A	*	T	G	T
H11	Clade I	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

TABLE 5 GenBank accession numbers of the 60 donkey mtDNA D-loop sequences used in the present study

Country	Number	GenBank accession nos.	Reference
Egypt	15	MG656081.1~MG656095.1	(Xia et al., 2019)
Ethiopia	15	MG656111.1~MG656125.1	(Xia et al., 2019)

Country	Number	GenBank accession nos.	Reference
Turkey	30	MH683672.1˜MH683701.1	(Ozkan Unal et al., 2020)

FIGURE 1 The geographic location of the Huaibei Grey donkey

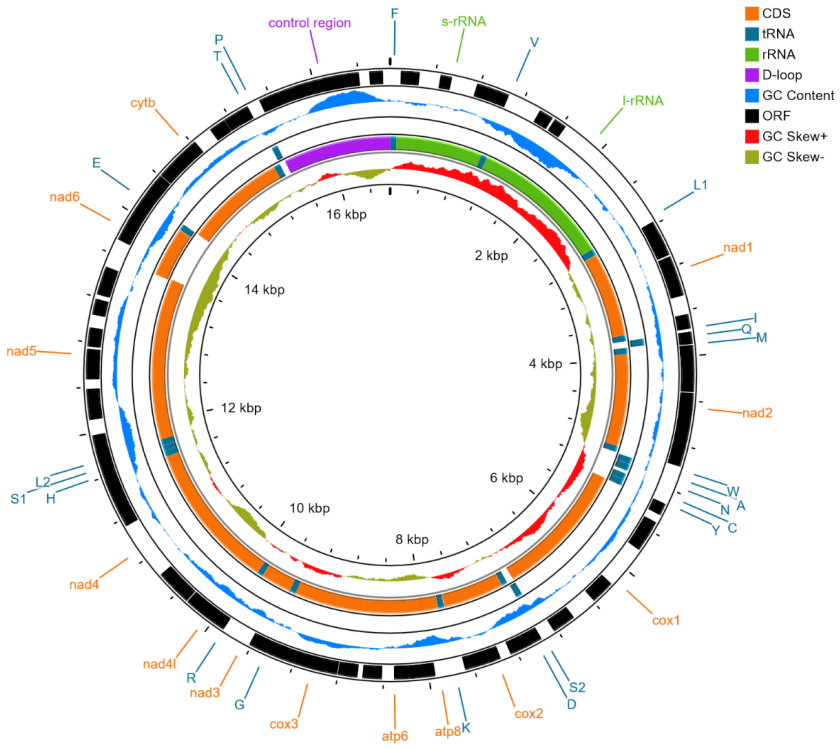

FIGURE 2 Gene map of the Huaibei Grey donkey

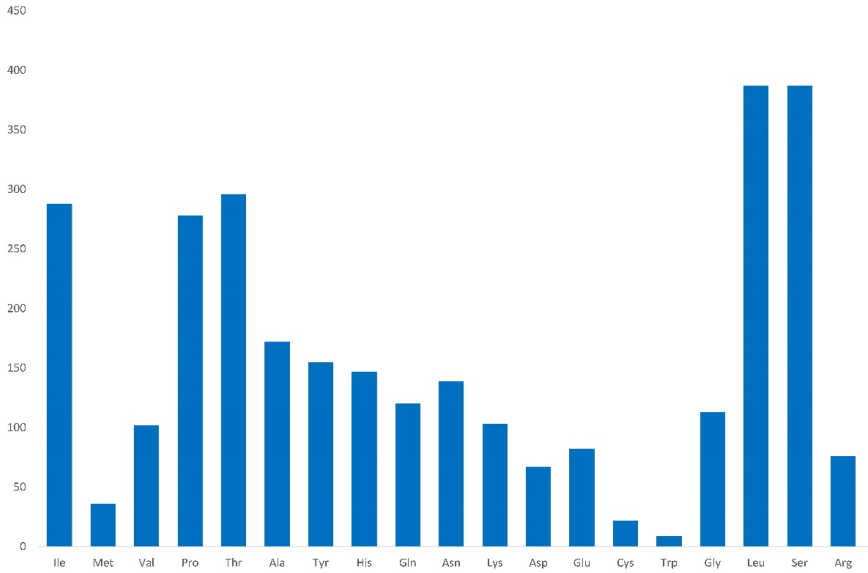

FIGURE 3 Codon distribution of the Huaibei Grey donkey mitogenome

The Y- and X-axis refer to the total number of codons and codon families, respectively.

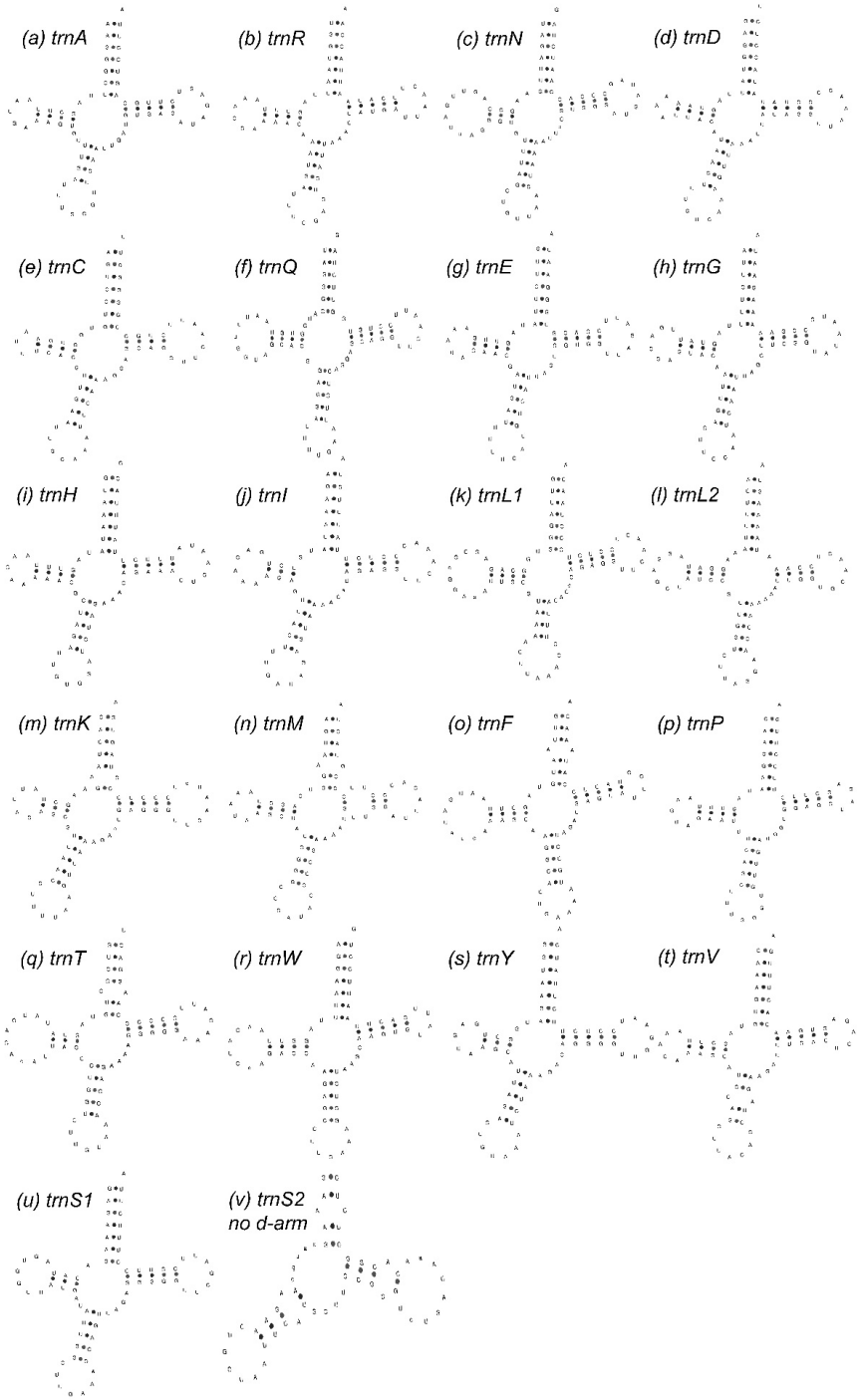

FIGURE 4 Putative tRNA secondary structures predicted from the 22 tRNA gene sequences of the Huaibei Grey donkey mitogenome

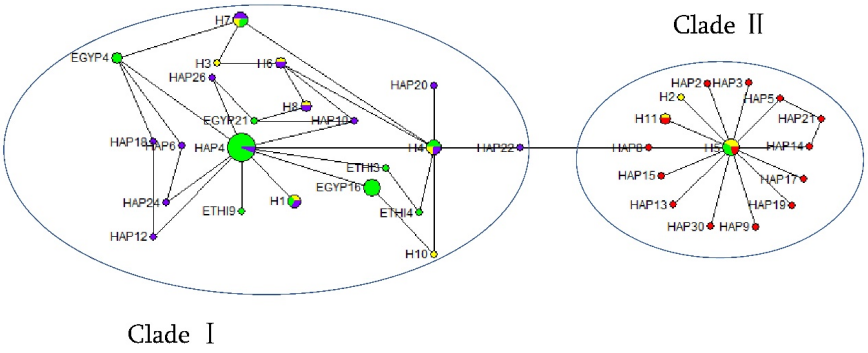

FIGURE 5 Median-joining network constructed from 11 haplotypes obtained from Huaibei grey donkey and 60 reference sequences obtained from 5 different populations

The circle areas correspond to the haplotype frequency. Yellow circles are haplotypes in this study. Green circles are 30 haplotypes which were found previously (Xia et al., 2019). The reference haplotypes Hap1˜Hap30 were downloaded from NCBI (Ozkan Unal et al., 2020), purple circles are haplotypes include Hap4、Hap6、Hap7、Hap10、Hap12、Hap16、Hap18、Hap20、Hap22、Hap23、Hap24、Hap25、Hap26、Hap27and Hap29, which belong to Clade I; purple circles are other haplotypes (Clade II).

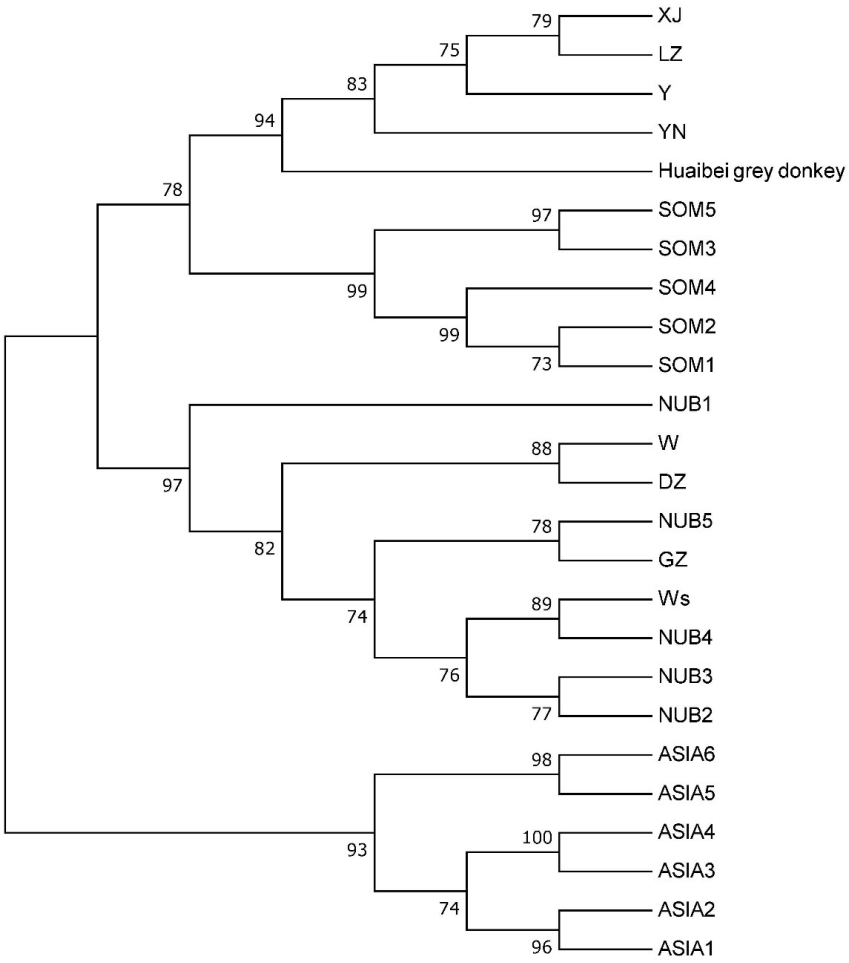

FIGURE 6 Phylogenetic relationships of Huaibei grey donkey determined using concatenated nucleotide sequences

Hosted file

`Tables-huaibei grey donkey.docx` available at https://authorea.com/users/501296/articles/581863-the-complete-mtdna-genome-of-huaibei-grey-donkey-genome-characterization-and-phylogenetic-analysis

Hosted file

`Figures-huaibei grey donkey.docx` available at https://authorea.com/users/501296/articles/581863-the-complete-mtdna-genome-of-huaibei-grey-donkey-genome-characterization-and-phylogenetic-analysis